A small energy reward was also provided when the secondary structure of the query chain was consistent with the template structure. For all residues that were in extended or helical states (as defined in the loose conformational definition used for the generic short range potentials) and that were in agreement with the secondary structure read from the corresponding fragments of the template protein, the system was stabilized by an energy equal to $-\varepsilon_{gen}$.

With the above restraints, the system only paid a small energetic penalty for moving along the template tube (shifts in the alignment with possible lateral adjustment); however, the penalty was large for escaping from the loosely defined volume occupied by the template. For instance, it was possible that continuous fragments of the original alignments permute (this cannot be called an alignment in the conventional sense) by swapping their original tube compartments. This only occurred when the potential strongly favored such a rearrangement of the topology. The two assignments, carried out by the algorithm, played a different role. The "old" one bonded the model chain to the broad vicinity of the threading-based template. The "new" dynamic assignment was a compromise between the template restraints and packing requirements of the model chain.

Summary of the threading model refinement protocol

The entire model building procedure is illustrated in a flow-chart (Figure 15) and can be outlined as follows:

a) generate the threading alignment between the query sequence and the template structure;

b) derive the sequence similarity based short and long-range pairwise potentials. The structures of proteins homologous to the query sequence are excised from the structural database; however, multiple alignments with the homologous sequences of unknown structures were used in the potential derivation procedures;

c) build the starting continuous model chain onto the lattice projected template structure;

d) build the tube around the aligned fragments of the template structure. Then, perform the first state of Monte Carlo refinement, where simulated annealing is done over a temperature range of 2-1. Since the Monte Carlo algorithm corrects unlike fragments of the alignment, the simulated annealing run is repeated two times. Subsequent runs have no systematic effect on the obtained models;

e) refinement of the structure. The model obtained from the above simulations is assumed to be the new template, with a full length, complete self-alignment. The distance restraints from the new template are narrowed to 4 lattice units, and simulated annealing is performed over a narrower temperature range (1.5 to 1.0);

f) selection of the lowest energy structures, by short isothermal simulations at T=1, followed by building the all-atom models using MODELLER.[24]

Results

Test proteins, templates and starting alignments

Twelve pairs of target/template proteins of very low sequence similarity were selected for the present study. These proteins belong to various classes of small globular proteins, with the selected set being rather representative. As described in the Methods section, the relative scaling of the various potentials of the model force field has been adjusted in a series of *ab initio* folding simulations on several (different from described here) small proteins. For the tuning of the template restraint contribution, three proteins, 2pcy, 256b and lhom, were selected. These proteins belong to rather different structural classes: 2pcy is a quite irregular β-type protein with a very poor initial threading-based model, when the 2azaA template is used. 256b is a compact, four-helix bundle, where the original alignment appears to

be quite good; however, the template and target structures have a different packing of helices that needs to be significantly readjusted to obtain a reasonable model. A very different example is 1hom. Here, the target fold is not very compact, and it is important to see if the proposed procedure can handle such small open structures. All proteins were subject to the previously described model building/refinement procedure. The list of these proteins is given in Table VIII. The threading alignments have been generated by a standard threading algorithm.[15] These alignments are compiled in Table IX. Tables VIII and IX appear below.

Table VIII. List of target/template pairs studied in this work

Target Protein			**Template Protein**		
PDB Code	**Name**	**Length**	**PDB Code**	**Name**	**Length**
laba_	Glutaredoxin	87	lego_	Glutaredoxin	85
1bbhA	Cytochrome C	131	2ccy_	Cytochrome C	127
1cewI	Cystatin	108	1molA	Monellin	94
1hom_	Antennapedia protein	68	11fb_	Transcription factor	77
1stfI	Papain	98	1molA	Monellin	94
1tlk_	Telokin	103	2rhe_	Immunoglobulin	114
256bA	Cytochrome C	106	1bbh_	Cytochrome C	131
2azaA	Azurin	129	1paz_	Pseudoazurin	120
2pcy_	Plastocyanin	99	2azaA	Azurin	129
2sarA	Ribonuclease	96	9rnt_	Ribonuclease	104
3cd4_	T-cell surface glycoprotein	178	2rhe_	Immunoglobulin	114
5fdl_	Ferrodoxin	106	2fxd_	Ferrodoxin	81